УДК: 004.032.26; 004.93; 004.93'1 https://doi.org/10.15407/jai2024.04.115

В. І. Слюсар
Центральний науково-дослідний інститут озброєння та військової техніки Збройних Сил України, Україна
пр. Повітрофлотський, 28Б, м. Київ, 03049
Інститут проблем штучного інтелекту МОН України і НАН України, Україна
пр. Академіка Глушкова, 40, Київ, 03680
https://orcid.org/0000-0002-2912-3149

ІНВЕРСНА АРХІТЕКТУРА U-NET – INVU-NET

V. Slyusar
Central Research Institute of Armaments and Military Equipment of Armed Forces of Ukraine, Ukraine
28B, Povytrophlotsky Ave, Kyiv, 03049
Institute of Artificial Intelligence Problems of the Ministry of Education and Science of Ukraine
and the National Academy of Sciences of Ukraine, Ukraine
40, Akademika Glushkova Ave., Kyiv, 03680
https://orcid.org/0000-0002-2912-3149

INVERSE ARCHITECTURE U-NET – INVU-NET

Анотація. У статті запропоновано інверсну архітектуру нейронної мережі U-Net, названу InvU-Net, яка відрізняється від традиційної схеми збільшенням розмірності зображень на початкових етапах обробки. Проведено порівняння двох підходів до збільшення роздільної здатності зображень: шарів UpSampling2D і транспонованих згорткових шарів Conv2DTranspose. Останні показали кращі результати завдяки можливості навчання вагових коефіцієнтів. У межах дослідження створено і протестовано кілька модифікацій InvU-Net: Small, Medium і Large, які відрізняються складністю структури, кількістю шарів і параметрів. Для підвищення точності сегментації було запропоновано інтеграцію механізмів уваги з метою покращення релевантності обробки ознак. Експерименти виявили, що спрощення механізмів уваги, включаючи зменшення кількості параметрів і вибір оптимальних точок інтеграції, дозволяє досягати високих результатів за меншої обчислювальної складності. Найкраща розроблена модель зі спрощеним механізмом уваги досягла точності 95.6%, перевершивши більш масштабні архітектури. Отримані результати демонструють перспективність використання InvU-Net для завдань сегментації та пропонують подальші напрямки оптимізації, зокрема використання адаптивних механізмів уваги та автоматизацію вибору параметрів нейромережі.
Ключові слова: сегментація зображень, нейронні мережі, архітектура U-Net, InvU-Net, згорткові шари, Conv2DTranspose, UpSampling2D, механізм уваги, оптимізація параметрів.

Abstract. The article proposes an inverse architecture of the U-Net neural network, named InvU-Net, which differs from the traditional scheme by increasing the dimensionality of images during the initial stages of processing. A comparison was conducted between two approaches for increasing image resolution: UpSampling2D layers and transposed Conv2DTranspose convolutional layers. The latter demonstrated superior results due to its ability to learn weighting coefficients. As part of the study, several InvU-Net modifications were developed and tested: Small, Medium, and Large, differing in structural complexity, the number of layers, and parameters. To improve segmentation accuracy, the integration of attention mechanisms was proposed to enhance the relevance of feature processing. Experiments revealed that simplifying attention mechanisms, including reducing the number of parameters and optimizing integration points, achieves high performance with lower computational complexity. The best-performing model, which incorporated a simplified attention mechanism, achieved 95.6% accuracy, surpassing larger architectures. The results highlight the potential of InvU-Net for segmentation tasks and suggest further optimization directions, such as employing adaptive attention mechanisms and automating the selection of neural network parameters.
Keywords: image segmentation, neural networks, U-Net architecture, InvU-Net, convolutional layers, Conv2DTranspose, UpSampling2D, attention mechanism, parameter optimization.

В задачах сегментації і класифікації зображень [1 - 3] широке розповсюдження отримали нейронні мережі типу U-Net [3 - 13]. Вони представляють собою подальший розвиток архітектури автоенкодера, в якій початкове зображення поступово трансформується у сцену малого розміру в так званому латентному просторі, а потім знову відновлюється в первинному форматі. При цьому в U-Net використовується передача масивів даних із сегмента зменшувальної розмірності до відповідного за розміром шару сегмента збільшення розмірності зображень.

Разом з тим, окрім такої архітектури можливе використання інверсної структури

UNet, в якій первісне зображення замість первинного зменшення розмірності піддається навпаки його збільшенню, а потім після проходження шару із найбільшим форматом знову зменшується до відновлення первісної розмірності на виході нейронної мережі. Недоліком такої структури є більший обсяг апаратних ресурсів, необхідних для її реалізації. Однак, у разі, якщо ресурсні витрати не вважаються критичними, цей підхід заслуговує уваги як альтернатива традиційним рішенням і цілком має право на існування.

Для підтвердження працездатності інверсної структури UNet (Inverce U-Net, InvU-Net) проведемо її навчання щодо завдання сегментації зображень літаків на тлі неба. При цьому використаємо датасет, раніше задіяний в [2] (рис. 1).



Рис. 1. Приклади зображень з датасету сегментації зображень літаків

На першому етапі розглянемо структуру InvU-Net small. В якості шарів, що реалізують збільшення формату зображення, використаємо шари Conv2DTranspose у поєднанні з шаром BatchNormalization [3]. Слід зазначити, що окрім Conv2DTranspose для збільшення розмірності даних може бути також застосована операція UpSampling2D. Однак порівняння цих двох альтернативних підходів дозволяє зробити наступні висновки.

UpSampling2D, як відомо, є шаром з бібліотеки Keras, який призначений для збільшення розмірів вхідних даних шляхом повторення їх значень або інтерполяції, залежно від заданого методу налаштування (nearest або bilinear). Зокрема, при використанні параметра nearest UpSampling2D повторює значення кодів сусідніх пікселів, а при виборі білінійної інтерполяції (bilinear) формуються проміжні значення між кодами пікселів відповідного оточення для більш плавних переходів між збільшеними сегментами зображення. Важливо, що шар UpSampling2D зовсім не має вагових коефіцієнтів і не навчається. Він просто збільшує зображення на основі визначеного розміру (наприклад, у 2 рази) без додаткових обчислень. Тому цей метод може призвести до втрати просторової інформації і може ускладнити для моделі точну локалізацію ознак на зображенні. Як наслідок це не дозволяє мережі адаптивно налаштовуватися на збільшення роздільної здатності, хоча з точки зору обчислювальної складності відповідна процедура є доволі простою.

Альтернативні транспоновані згорткові шари Conv2DTranspose, також відомі як “деконволюційні” шари, працюють на основі оберненої операції звичайної згортки. Це означає, що фактично Conv2DTranspose намагається “розгортати” карти ознак і збільшувати їх просторову роздільну здатність без втрати інформації, використовуючи для цього

фільтри з ваговими коефіцієнтами. В результаті, під час навчання можуть бути підібрані оптимальні ваги для підвищення якості масштабування, що дозволить більш точно локалізувати ознаки зображення, зберігаючи його структуру для досягнення більш високої точності сегментації зображень.

Структура першої з досліджених автором нейронних мереж типу InvU-Net представлена на рис. 2. Для зручності подальшого порівняння результатів бажано кожному варіанту моделі нейронної мережі присвоювати своє умовне найменування, тому в даному випадку назвемо просту нейронну мережу як InvU-Net Small-1.

Детальний опис цієї структури наведено у табл. 1. В ній відображено послідовність шарів у нейронній мережі, їх типи, вихідні формати даних та кількість параметрів для кожного шару. Нумерація шарів у першому лівому стовпці не збігається з оригінальною нумерацією в моделі, оскільки їх пронумеровано послідовно для зручності читання таблиці.

Мережа починається з вхідного шару (InputLayer), який приймає зображення розміром 128×160 пікселів із 3-колірними каналами (RGB). Після вхідного шару йде серія згорткових блоків. Кожен блок складається з шару згортки (Conv2D) та наступного за ним шару пакетної нормалізації (BatchNormalization). Ця комбінація допомагає отримувати ознаки із зображення та стабілізувати навчання.

Шар Conv2DTranspose збільшує удвічі розмірність даних - з 128×160 до 256×320 пікселів, після чого за допомогою шару MaxPooling2D просторові розміри знову повертаються до формату 128×160.

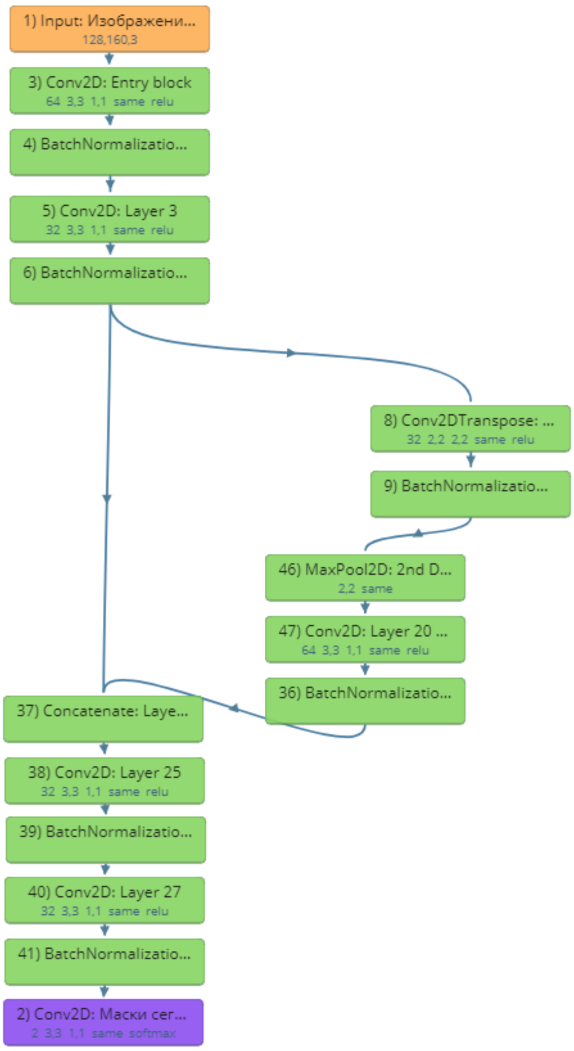


Рис. 2. Структура нейронної мережі InvU-Net Small-1

Таблиця 1. Складові елементи нейромережі InvU-Net Small-1

№	Шар	Тип	Вихідний формат	Кількість параметрів
1	1	InputLayer	(None, 128, 160, 3)	0
2	Conv2D_3	Conv2D	(None, 128, 160, 64)	1792
3	BatchNormalization_4	BatchNormalization	(None, 128, 160, 64)	256
4	Conv2D_5	Conv2D	(None, 128, 160, 32)	18464
5	BatchNormalization_6	BatchNormalization	(None, 128, 160, 32)	128
6	Conv2DTranspose_8	Conv2DTranspose	(None, 256, 320, 32)	4128
7	BatchNormalization_9	BatchNormalization	(None, 256, 320, 32)	128
8	MaxPool2D_46	MaxPooling2D	(None, 128, 160, 32)	0
9	Conv2D_47	Conv2D	(None, 128, 160, 64)	18496
10	BatchNormalization_36	BatchNormalization	(None, 128, 160, 64)	256
11	Concatenate_37	Concatenate	(None, 128, 160, 96)	0
12	Conv2D_38	Conv2D	(None, 128, 160, 32)	27680
13	BatchNormalization_39	BatchNormalization	(None, 128, 160, 32)	128
14	Conv2D_40	Conv2D	(None, 128, 160, 32)	9248
15	BatchNormalization_41	BatchNormalization	(None, 128, 160, 32)	128
16	2	Conv2D	(None, 128, 160, 2)	578

Шар Concatenate_37 поєднує виходи з попередніх шарів з поточним потоком даних. Це допомагає зберегти інформацію з різних рівнів абстракції та покращує градієнтний потік. Мережа закінчується кількома згортковими шарами, поступово зменшуючи кількість каналів. Останній шар (Conv2D) виводить тензор із 2 каналами, що відповідає завданню сегментації.

В цілому, нейронна мережа InvU-Net Small-1 характеризується досить малими розмірами, оскільки загальна кількість навчальних параметрів в ній становить 80898. Це дозволяє використовувати її на пристроях з обмеженими ресурсами для завдань сегментації, покращення роздільної здатності або виділення певних ознак на зображенні.

Для навчання описаної та усіх розглянутих далі моделей нейромереж використовувався платний сервіс Google Colab, який передбачав оренду графічної карти (GPU) A100-SXM4-40GB з обсягом відеопам'яті VRAM 40 ГБ і розміром виділеного оперативного запам'ятовуючого пристрою (ОЗП) процесора 83,48 ГБ. Процес навчання проводився протягом 100 епох порціями по 20 епох із зупинками. При цьому середній час навчання протягом 20 епох склав всього 58 с. У пікових моментах завантаження пам'яті відеокарти досягало 85 %, а ОЗП процесора – 11,3 %. Розмір батчу був заданий 32, в якості оптимізатора вибраний алгоритм Adam. При цьому величина кроку зміни вагових коефіцієнтів Learning Rate=0,001 залишалася фіксованою протягом всього процесу навчання.

Найкращий досягнутий результат на навчальній вибірці склав 97 %, а на перевірочній – 93,9%. Відповідні графіки навчання представлені на рис. 3. В цілому отримані результати виявилися досить надихаючими і підтвердили працездатність нової архітектури U-Net, що дозволило перейти до її детального дослідження та вивчення впливу різноманітних покращень у структурі на досягнуту точність сегментації.



Рис. 3. Графіки навчання нейромережі InvU-Net Small-1

Подальші зміни в архітектуру нейромережі InvU-Net були внесені у напрямку її поступового ускладнення. Зокрема, більш розвинена архітектура малорозмірної InvU-Net представлена на рис. 4 під назвою InvU-Net Medium. Особливістю її є нарощування кількості шарів в структурі латентного простору, двоступеневе застосування операції Conv2DTranspose з підвищенням у ній кількості ядер від 64 до 128 та інші відповідні зміни. При цьому загальна кількість шарів у моделі сягнула 24, а максимальна розмірність масиву даних в латентному просторі збільшилася до величини 1024×1280×16. Внаслідок таких змін загальна кількість навчених параметрів в моделі зросла більше ніж в 3 рази і становила 274338. Їхній розподіл по шарах архітектури нейромережі представлено в табл. 2.

Для збереження спадковості параметри навчального процесу нейромережі були залишені без змін. Внаслідок навчання представленої на рис. 4 архітектури вдалося отримати підвищення точності сегментації зображень на перевірочній вибірці до 94.7%.

Для здійснення переходу до більш масштабованої порівняно з рис. 4 архітектури можна виконати кілька змін та розширень.

По-перше, варто додати більше паралельних гілок, що дозволить посилити гнучкість обробки різних рівнів характеристик зображень. Це досягається шляхом додавання шарів Concatenate, які об'єднують інформацію з різних сегментів, створюючи складнішу модель. На кожному рівні обробки додаються додаткові шляхи, що допомагають моделі вивчати ілюстративно багатші представлення ознак. Додатково, у масштабованій версії після кожного об’єднувального шару слід додати нові шари Conv2D і BatchNormalization, що підвищить здатність мережі виявляти та обробляти детальніші характеристики, виділені на попередньому рівні. Це робить пропоновану архітектуру більш глибокою і потужною для вирішення складних задач. Загальний вид удосконаленої структури нейромережі наведено на рис. 5, а детальний опис параметрів її шарів - у таблиці 3.

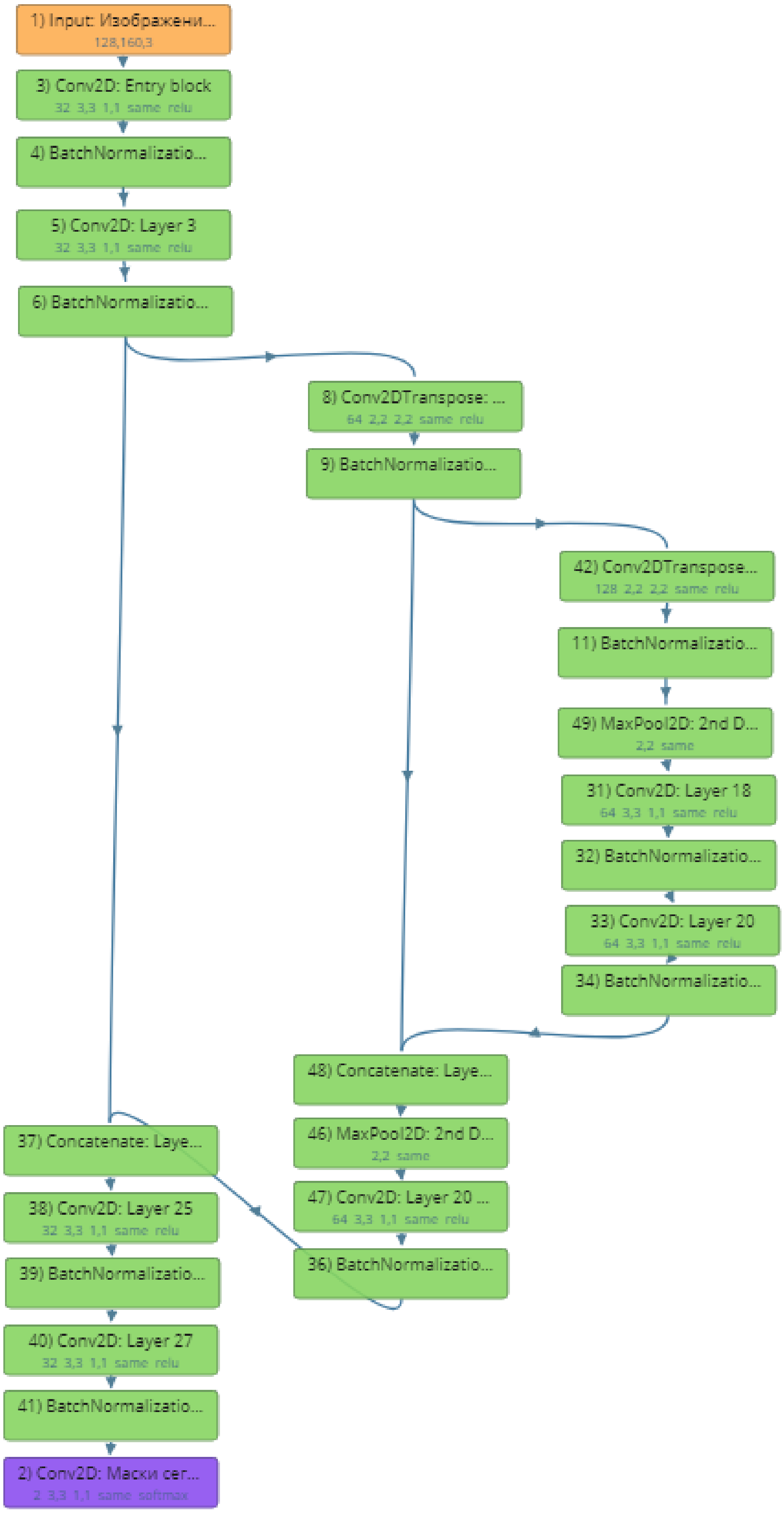


Рис. 4. Нейромережа InvU-Net Medium

Крім вже описаних змін у структурі InvU-Net Large, у її масштабованій архітектурі збільшено використання шарів типу MaxPooling для зменшення розмірів масивів даних, що дозволяє виділяти більш абстрактні та узагальнені характеристики на глибших рівнях. На шляху до латентного простору використовуються 3 шари Conv2DTranspose, що дозволяє масштабувати ознаки на більші розміри тензорів даних для отримання збагаченої інформації. Кількість фільтрів у шарах Conv2D також збільшена, що дозволяє масштабованій архітектурі аналізувати більше представлень масиву даних на кожному рівні, допомагаючи моделі краще налаштовуватися на складні завдання сегментації зображень за рахунок додаткової обчислювальної потужності.

Таблиця 2. Складові елементи нейромережі InvU-Net Medium

№	Шар	Тип	Вихідний формат	Кількість параметрів
1	1	InputLayer	(None, 128, 160, 3)	0
2	Conv2D_3	Conv2D	(None, 128, 160, 32)	896
3	BatchNormalization_4	BatchNormalization	(None, 128, 160, 32)	128
4	Conv2D_5	Conv2D	(None, 128, 160, 32)	9248
5	BatchNormalization_6	BatchNormalization	(None, 128, 160, 32)	128
6	Conv2DTranspose_8	Conv2DTranspose	(None, 256, 320, 64)	8256
7	BatchNormalization_9	BatchNormalization	(None, 256, 320, 64)	256
8	Conv2DTranspose_42	Conv2DTranspose	(None, 512, 640, 128)	32896
9	BatchNormalization_11	BatchNormalization	(None, 512, 640, 128)	512
10	MaxPool2D_49	MaxPooling2D	(None, 256, 320, 128)	0
11	Conv2D_31	Conv2D	(None, 256, 320, 64)	73792
12	BatchNormalization_32	BatchNormalization	(None, 256, 320, 64)	256
13	Conv2D_33	Conv2D	(None, 256, 320, 64)	36928
14	BatchNormalization_34	BatchNormalization	(None, 256, 320, 64)	256
15	Concatenate_48	Concatenate	(None, 256, 320, 128)	0
16	MaxPool2D_46	MaxPooling2D	(None, 128, 160, 128)	0
17	Conv2D_47	Conv2D	(None, 128, 160, 64)	73792
18	BatchNormalization_36	BatchNormalization	(None, 128, 160, 64)	256
19	Concatenate_37	Concatenate	(None, 128, 160, 96)	0
20	Conv2D_38	Conv2D	(None, 128, 160, 32)	27680
21	BatchNormalization_39	BatchNormalization	(None, 128, 160, 32)	128
22	Conv2D_40	Conv2D	(None, 128, 160, 32)	9248
23	BatchNormalization_41	BatchNormalization	(None, 128, 160, 32)	128
24	2	Conv2D	(None, 128, 160, 2)	578

Додані на рис. 5 шари Concatenate поєднують інформацію з кількох рівнів обробки, що забезпечує передачу інформації між різними рівнями та гілками, дозволяючи уникнути втрати деталей при переході між процесами зменшення та збільшення розмірів. Такий підхід сприяє підвищенню якості обробки, оскільки різні характеристики комбінуються і взаємодіють. При цьому кількість параметрів, доступних для навчання, зросла до 8894786. В результаті, цілком очікувано, вказані удосконалення дозволили досягти точності сегментації 95.2 %. Отже, представлені результати підтвердили правильність обраного шляху удосконалення запропонованої архітектури.

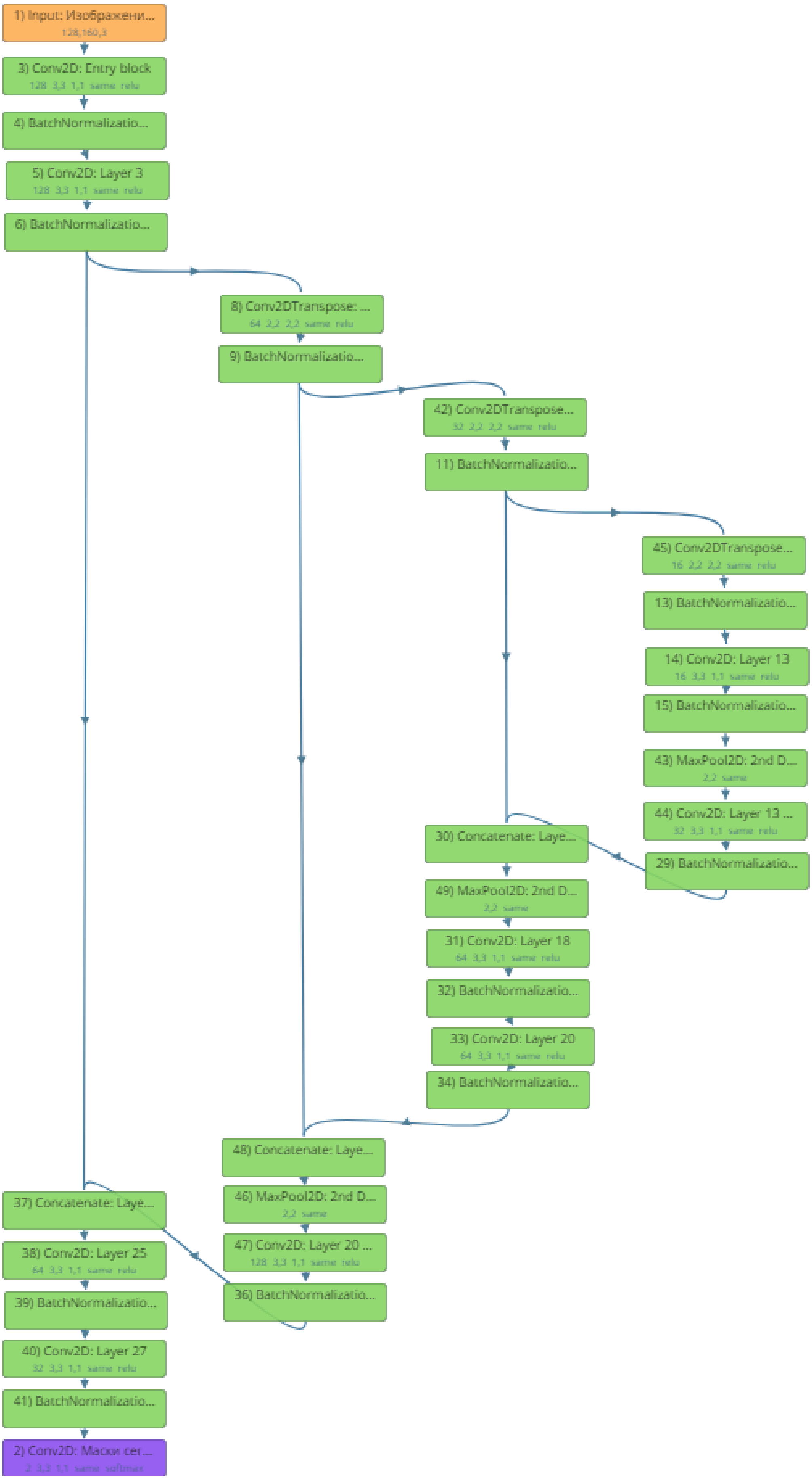


Рис. 5. Ускладнена архітектура InvU-Net з умовною назвою InvU-Net Large

Стосовно досягнутого рівня точності слід вказати, що для її підвищення потрібно оптимізувати параметри шарів, наприклад, за допомогою генетичного алгоритму або AutoKeras, однак їх розгляд виходить за межі даної статті.

Таблиця 3. Складові елементи нейромережі InvU-Net Large

№	Шар	Тип	Вихідний формат	Кількість параметрів
1	1	InputLayer	(None, 128, 160, 3)	0
2	Conv2D_3	Conv2D	(None, 128, 160, 128)	3584
3	BatchNormalization_4	BatchNormalization	(None, 128, 160, 128)	512
4	Conv2D_5	Conv2D	(None, 128, 160, 128)	147584
5	BatchNormalization_6	BatchNormalization	(None, 128, 160, 128)	512
6	Conv2DTranspose_8	Conv2DTranspose	(None, 256, 320, 64)	32832
7	BatchNormalization_9	BatchNormalization	(None, 256, 320, 64)	256
8	Conv2DTranspose_42	Conv2DTranspose	(None, 512, 640, 32)	8224
9	BatchNormalization_11	BatchNormalization	(None, 512, 640, 32)	128
10	Conv2DTranspose_45	Conv2DTranspose	(None, 1024, 1280, 16)	2064
11	BatchNormalization_13	BatchNormalization	(None, 1024, 1280, 16)	64
12	Conv2D_14	Conv2D	(None, 1024, 1280, 16)	2320
13	BatchNormalization_15	BatchNormalization	(None, 1024, 1280, 16)	64
14	MaxPool2D_43	MaxPooling2D	(None, 512, 640, 16)	0
15	Conv2D_44	Conv2D	(None, 512, 640, 32)	4640
16	BatchNormalization_29	BatchNormalization	(None, 512, 640, 32)	128
17	Concatenate_30	Concatenate	(None, 512, 640, 64)	0
18	MaxPool2D_49	MaxPooling2D	(None, 256, 320, 64)	0
19	Conv2D_31	Conv2D	(None, 256, 320, 64)	36928
20	BatchNormalization_32	BatchNormalization	(None, 256, 320, 64)	256
21	Conv2D_33	Conv2D	(None, 256, 320, 64)	36928
22	BatchNormalization_34	BatchNormalization	(None, 256, 320, 64)	256
23	Concatenate_48	Concatenate	(None, 256, 320, 128)	0
24	MaxPool2D_46	MaxPooling2D	(None, 128, 160, 128)	0
25	Conv2D_47	Conv2D	(None, 128, 160, 128)	147584
26	BatchNormalization_36	BatchNormalization	(None, 128, 160, 128)	512
27	Concatenate_37	Concatenate	(None, 128, 160, 256)	0
28	Conv2D_38	Conv2D	(None, 128, 160, 3264)	7523520
29	BatchNormalization_39	BatchNormalization	(None, 128, 160, 3264)	13056
30	Conv2D_40	Conv2D	(None, 128, 160, 32)	940064
31	BatchNormalization_41	BatchNormalization	(None, 128, 160, 32)	128
32	2	Conv2D	(None, 128, 160, 2)	578

Більш пріоритетним завданням є дослідження впливу різноманітних механізмів уваги [14, 15], додавання яких до нейромережі для сегментації зображень сприймається як логічний крок з метою покращення її точності. Проте саме по собі це не гарантує автоматичного підвищення продуктивності моделі і може не призвести до очікуваних результатів, оскільки механізм уваги додає додаткові параметри та обчислювальну складність. Зокрема, якщо модель стає занадто складною для даного набору даних, це може призвести до її перенавчання або труднощів із збіжністю (конвергенцією) під час тренування. Як відомо [16], під конвергенцією розуміють процес поступового зменшення помилок моделі нейромережі до такого рівня, коли вона досягає стабільного стану, тобто коли її продуктивність на тренувальних і тестових даних перестає значно покращуватись. Це означає, що алгоритм навчання наближається до оптимального значення цільової функції (наприклад, мінімізації втрат), і вагові коефіцієнти моделі стають близькими до тих значень, при яких вона може найкраще виконувати завдання. Конвергенція зазвичай

оцінюється шляхом відстеження змін втрат (loss function) або інших метрик під час кожної епохи навчання. Коли ці значення перестають помітно змінюватися, вважається, що модель досягла конвергенції.

Також суттєво, щоб механізм уваги був правильно інтегрований до архітектури нейромережі, оскільки неправильне його розміщення або налаштування може порушити потік інформації та негативно вплинути на навчання моделі. Слід зазначити, що механізм уваги має свої власні гіперпараметри, які потребують належної оптимізації. Без ретельного налаштування ці параметри можуть не відповідати специфіці задачі, що знижує ефективність моделі. Крім того, ефективність механізму уваги може сильно залежати від характеристик вхідних даних. Зокрема, якщо дані не містять складних залежностей або контексту, який потребує уваги, додавання цього механізму може не дати переваг.

Збільшення обчислювальної складності за рахунок інтеграції механізму уваги може вимагати більше ресурсів для навчання та інференсу. У деяких випадках це може бути непрактичним або економічно невигідним. Тим більше, що коли в моделі вже використовуються методи нормалізації або регуляризації, механізм уваги може не дати додаткового покращення або навіть конфліктувати з ними.

Таким чином, при додаванні механізму уваги для підвищення точності нейромережі з сегментації зображень важливо проводити експерименти та аналізувати, чи дійсно цей механізм приносить користь у конкретному контексті. Ефективність моделі залежить від комплексного підходу, який включає правильний вибір архітектури, налаштування гіперпараметрів та розуміння природи даних і задачі.

З урахуванням зазначеного, замість класичних схем реалізації механізму уваги в нейромережі [14], розглянемо альтернативні підходи, що можуть сприяти підвищенню ефективності нейромережі й були вже апробовані в попередніх роботах автора [15, 17]. В якості такого альтернативного варіанта пропонується використати маскові механізми уваги, які можуть забезпечити більшу гнучкість та ефективність в порівнянні з традиційними підходами.

Як відомо, маскові механізми уваги не потребують збереження коефіцієнтів Q, V та K, тому що вони застосовують маски для визначення релевантності інформації, тим самим зменшуючи обсяг пам'яті, необхідної для обробки даних. При цьому для перемноження маски на зображення використовується поелементний добуток Адамара [15], який передбачає помноження відповідних елементів двох матриць однакових розмірів. У контексті маскових механізмів уваги ця операція дозволяє ефективно застосовувати маски до матриці уваги, відфільтровуючи нерелевантну інформацію. У більш загальному випадку може використовуватись проникаючий торцевий добуток, запропонований автором [18, 19] для більш глибокої інтеграції маскових механізмів уваги з іншими компонентами нейронної мережі, що дозволить підвищити її ефективність [15].

Як і передбачалося, спроби застосування маскового механізму уваги в розглянутому вище варіанті InvU-Net Small (рис. 1) не дозволили одразу досягти бажаного ефекту. Серед таких, не досить вдалих, варіантів слід послатися на архітектуру, показану на рис. 6, яка дозволила досягти точності 89.8 %. Такий результат, скоріш за все, пояснюється кількома переліченими вище чинниками: надмірним ускладненням нейромережі, кількість налаштовуваних параметрів в якій становить 838983075; обраною точкою підключення механізму уваги (до операції Conv2DTranspose) та його структурою.

Дещо кращий результат - 91.4 % точності сегментації, дозволили отримати удосконалення даної архітектури, представлені на рис. 7. Відмінності між двома моделями нейромереж (рис. 6 та 7) полягають, перш за все, у структурі механізму уваги. У першій моделі після шару Reshape_53 використано шар Conv2D_51 типу Conv2D з розмірністю

вихідного тензору 128×160×32 та 320 параметрами. У другій моделі після Reshape_53 використовується шар Conv2DTranspose_51 типу Conv2DTranspose з вихідною розмірністю 256×320×32 та 160 параметрами. Таким чином, проведена заміна дозволила скоротити кількість налаштовуваних параметрів у відповідному шарі механізму уваги у 2 рази. Інша суттєва відмінність полягає у точці підключення шару Multiply, який у моделі на рис. 7 комбінує нормалізовані виходи двох шарів зворотної згортки (Conv2DTranspose_8 після BatchNormalization_6 та Conv2DTranspose_51 механізму уваги), що утворює інший шлях передачі даних і впливає на остаточний результат. В цілому ж, загальна кількість параметрів у першій моделі становить 838983651, а у другій моделі – 838983491, що на 160 параметрів менше.

В обох розглянутих моделях нейромереж (рис. 6, 7) механізм уваги інтегрований через шар множення (Multiply), який формує поелементний добуток двох наборів активацій. Це відповідає концепції канальної уваги, коли модель адаптивно зважує значимість кожного каналу, підсилюючи релевантні та пригнічуючи менш важливі ознаки. Шар Multiply в цій архітектурі застосовує множення до всього тензору даних, що фактично діє як увага, накладаючи ваги, які підсилюють або пригнічують певні області вхідних даних. У наведеному випадку реалізації така операція може розглядатися також як маскова увага, оскільки створюється маска, що застосовується до всього тензора для підкреслення важливих ознак і зменшення впливу менш важливих. Цей підхід дозволяє моделі ефективніше використовувати інформацію, яка має найбільшу значимість для розв'язання поставленого завдання. Слід зауважити, що такий механізм не є самоувагою у класичному розумінні, як це реалізовано в трансформерах, де кожен елемент вхідної послідовності порівнюється з усіма іншими. У підсумку його можна вважати різновидом просторово-канальної уваги, оскільки маска дозволяє реалізувати просторову увагу до всіх пікселів у зображенні, визначаючи, які регіони повинні отримати більше уваги на підставі значень вагових параметрів у масці.

Зафіксувавши проведені зміни в архітектурі InvU-Net як позитивний тренд, подальшу її модифікацію зосередимо на ще більшому зменшенні кількості гіперпараметрів механізму уваги, знизивши у 10 разів розмірність другого шару Dense (Dense_55), що передує операції Reshape. Відповідні зміни відображені на рис. 8.

Результатом зазначеної модифікації лише одного шару у механізмі уваги стало зменшення загальної кількості параметрів моделі на 754993152 до величини 83990339, з яких для налаштування доступні 83989763. Така зміна позитивно вплинула на швидкість навчання та зменшила вимоги до апаратного забезпечення. Основний же результат полягає в отриманому підвищенні точності навчання, яка досягла 95.3 %.

Таким чином, основним висновком з експериментів щодо дослідження архітектур InvU-Net Small з механізмами уваги став вибір в якості ключового напряму для поліпшення ефективності даного типу нейромережі максимального спрощення структури вказаного механізму. Звісно, таке спрощення має здійснюватися за умови, що воно не обмежить чутливість моделі до складних патернів в даних, у поєднанні з попутною оптимізацією вибору точок підключення входу й виходу механізму уваги.

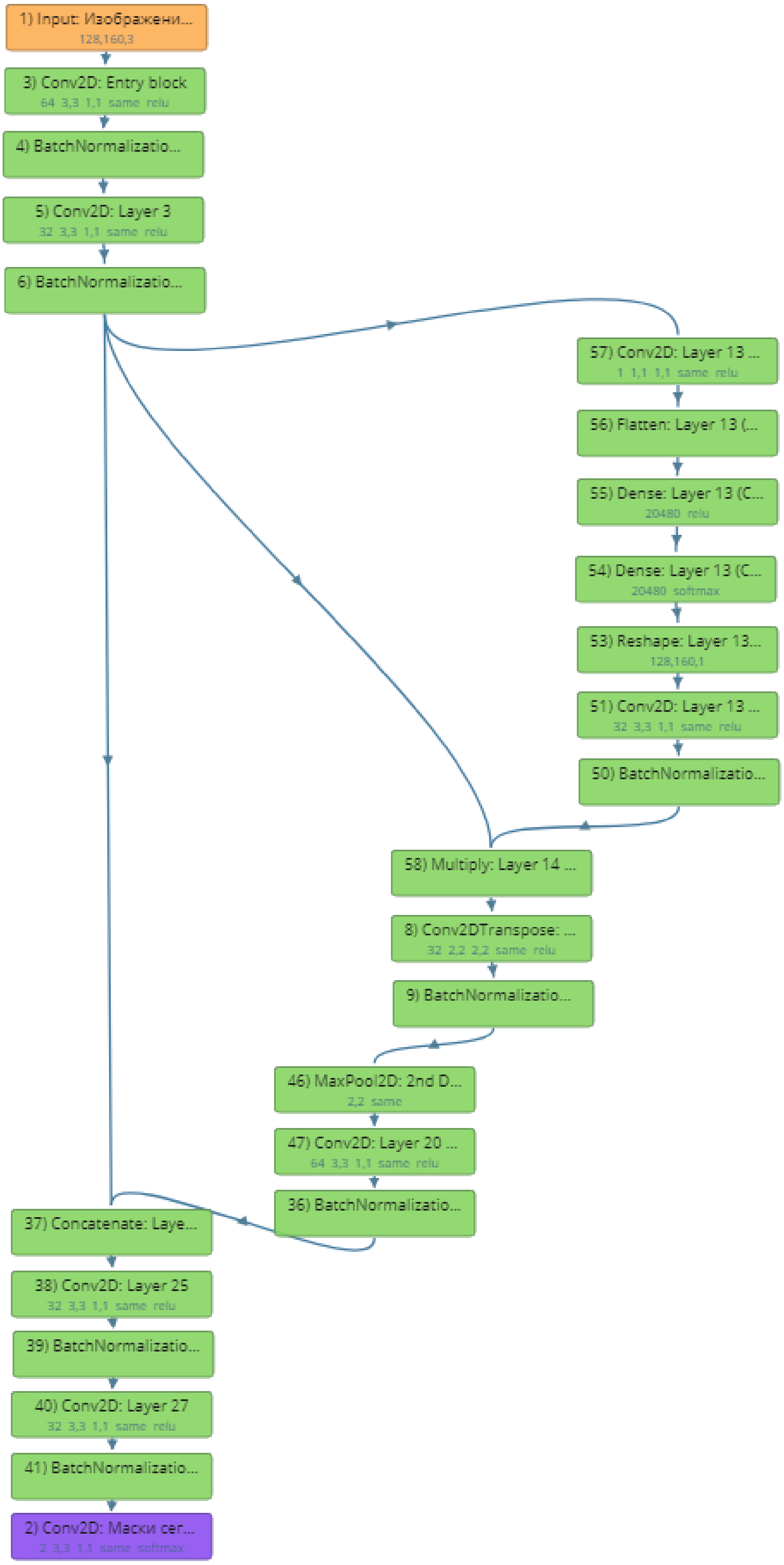


Рис. 6. Варіант InvU-Net з механізмом уваги на вході латентного простору

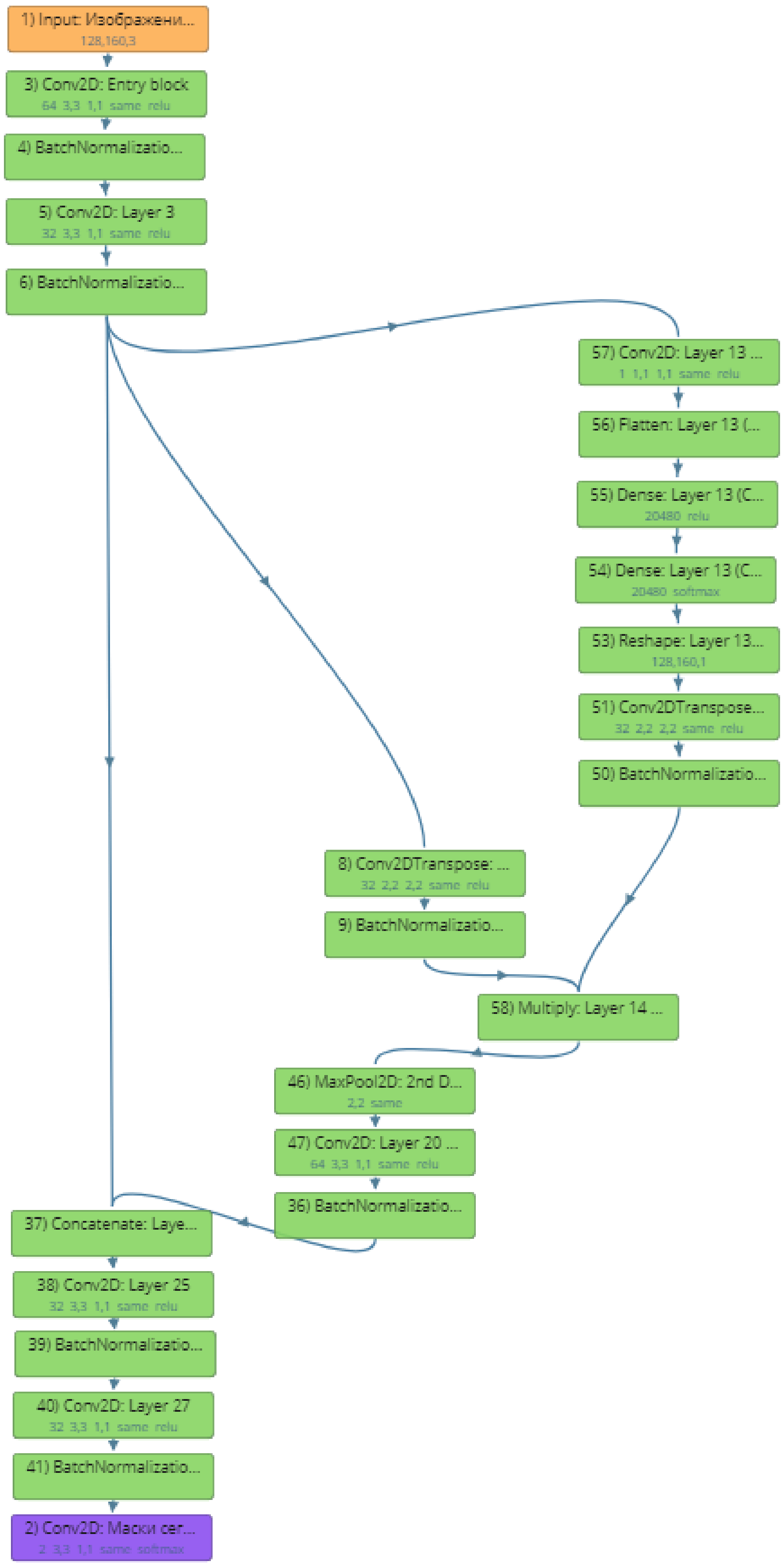


Рис. 7. Удосконалений варіант моделі, представленої на рис. 6

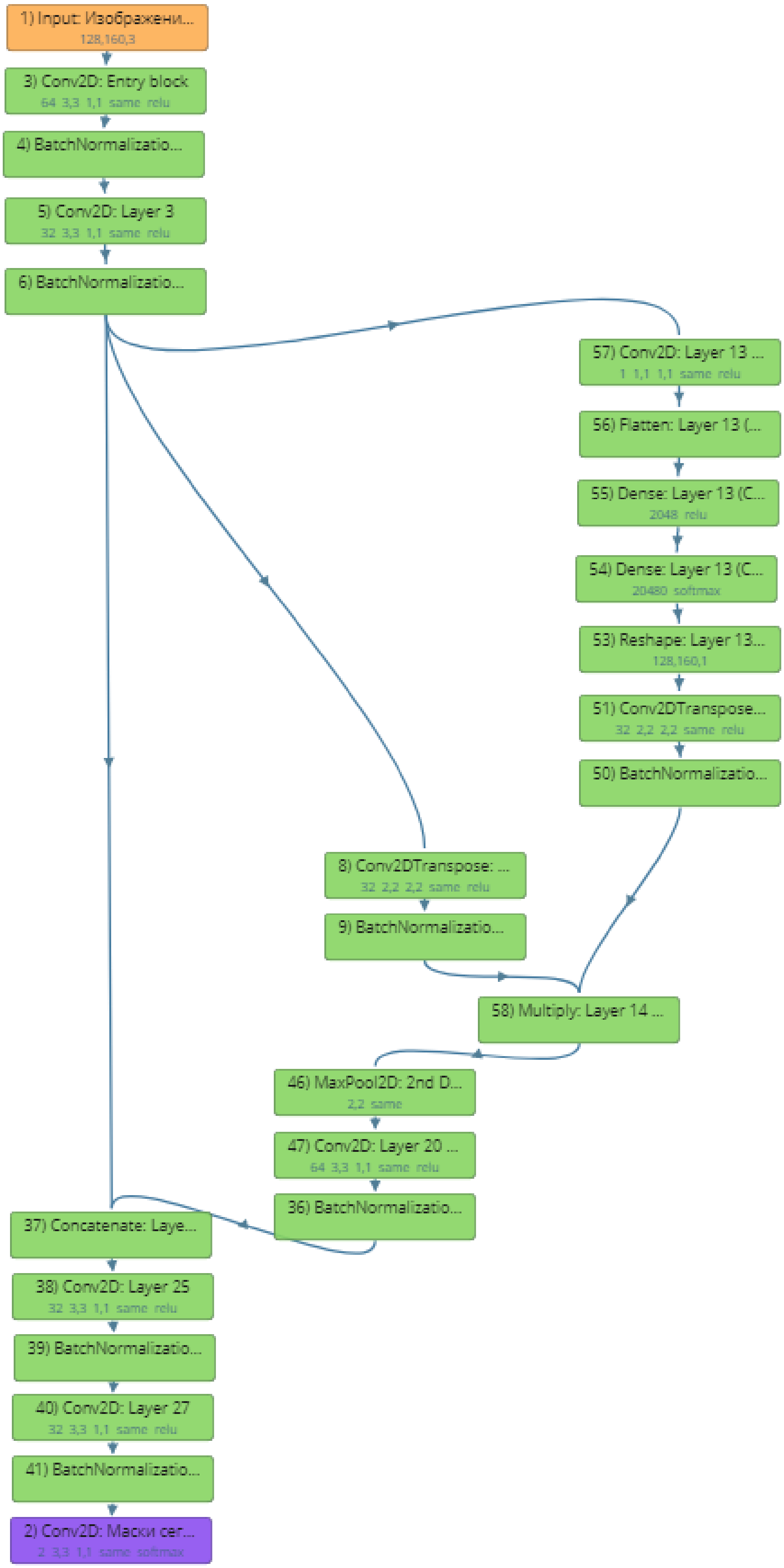


Рис. 8. InvU-Net зі зменшеною розмірністю шару Dense у механізмі уваги

Результатом подальшого пошуку стала нова структура InvU-Net Small, що проілюстрована на рис. 9.

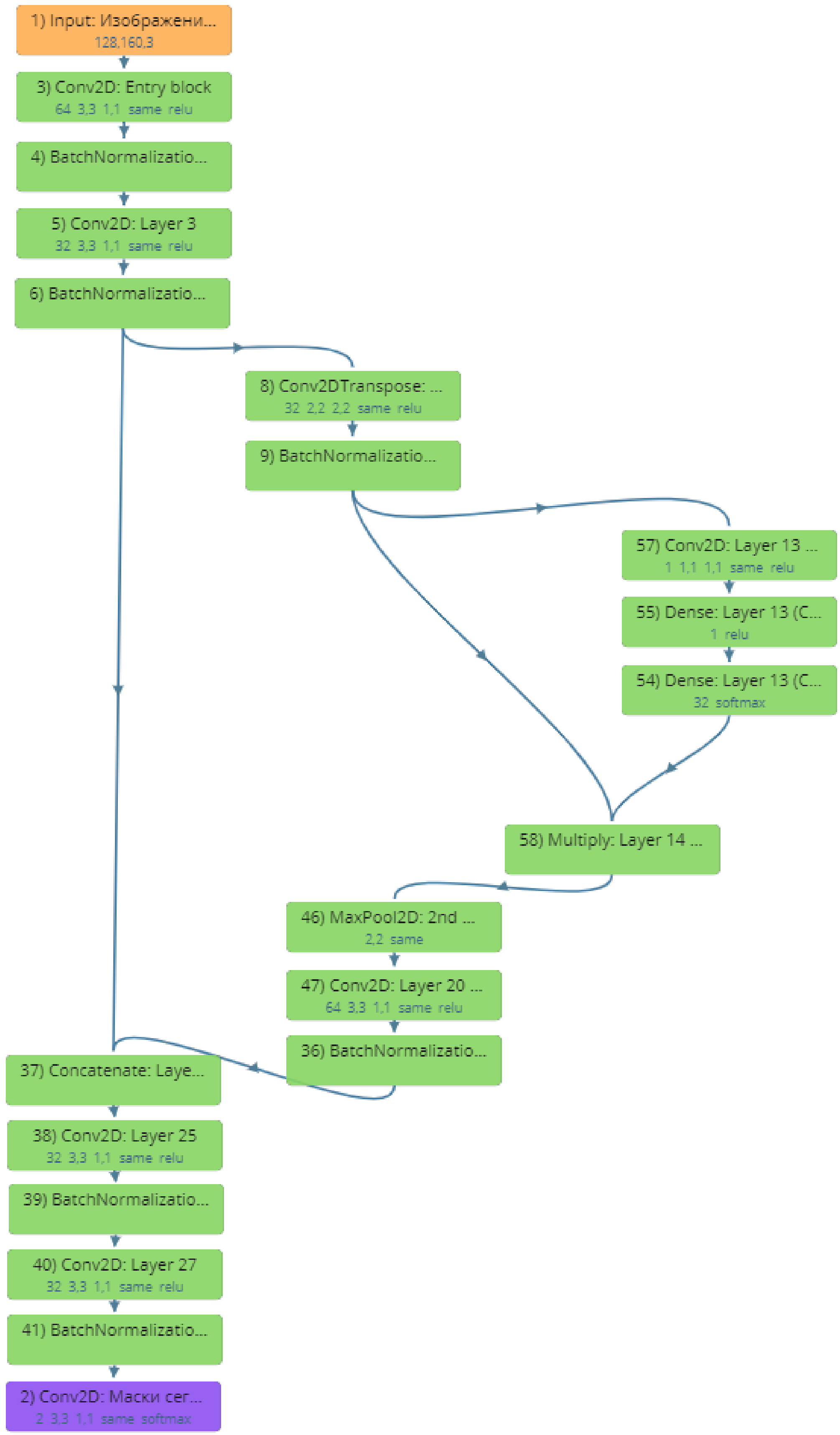


Рис. 9. InvU-Net зі спрощеним механізмом уваги

Особливістю архітектури (рис. 9) є тришаровий механізм уваги, що формально займає в ній місце латентного простору. На вході цього механізму підключений згортковий шар Conv2D, що слугує початковим етапом для вилучення базових ознак із вхідного тензора. Вилучені ознаки є фундаментальними для визначення тих областей вхідних даних, на які слід звернути більше уваги у наступних етапах механізму уваги. Фактично, згортковий шар виконує початкове перетворення просторових ознак, що готуються для подальшої обробки у механізмі уваги, а також генерує активації, які можуть використовуватися для створення контекстної інформації, необхідної для формування маски уваги. Крім того, Conv2D допомагає зменшити розмірність і виділити специфічні фільтри, що служать вхідними ознаками для наступних компонентів механізму уваги. На виходах шару Conv2D застосовуються нелінійні активаційні функції ReLU, які підвищують виразність моделі. Це особливо важливо в контексті уваги, оскільки нелінійні активації дозволяють моделі краще захопити складні локальні патерни, що робить створену маску уваги більш ефективною для посилення або пригнічення окремих частин вхідного тензора. Отримана в результаті контекстна інформація піддається подальшій обробці у повнозв'язних шарах Dense, щоб визначити релевантність кожного просторового елемента.

Таким чином, відмінність моделі на рис. 9 полягає у спрощеній структурі механізму уваги та його специфічному підключенні у середині латентного простору. Такі особливості дозволили отримати точність сегментації вказаною нейромережею на рівні 95.6 %. Досягнутий результат перевищує ефективність великої моделі InvU-Net Large, приведеної на рис. 5, і свідчить про результативність обраної стратегії удосконалення архітектури InvU-Net на основі застосування механізмів уваги.

Висновки

У ході дослідження розроблено і протестовано інверсну архітектуру U-Net, названу InvU-Net, використання якої є альтернативним підходом до вирішення завдань сегментації зображень.

Особливістю даного підходу є застосування збільшення розмірності зображення в межах енкодерного сегмента з подальшим його зменшенням в декодері до рівня вхідного масиву даних. Застосування транспонованих згорткових шарів (Conv2DTranspose) дозволило зберегти просторову інформацію та досягти адаптивного масштабування, що показало свою ефективність з точки зору точності сегментації. Запропоновані модифікації InvU-Net (Small, Medium, Large) продемонстрували гнучкість і можливість адаптації архітектури до різних рівнів ресурсів та вимог задач. Інтеграція механізмів уваги виявила їх потенціал для покращення точності сегментації, однак потребує ретельного налаштування та зменшення обчислювальної складності. При цьому досягнення точності 95.6% на спрощеній моделі InvU-Net Small зі скороченим механізмом уваги підтвердило ефективність обраного підходу до оптимізації.

Подальші дослідження можуть бути зосереджені на вивченні механізмів адаптивної уваги, автоматизованій оптимізації параметрів та застосуванні архітектури в задачах із високими вимогами до продуктивності й точності. Отримані результати відкривають нові перспективи у використанні інверсних архітектур у сфері комп’ютерного зору та мультимодальних великих мовних моделях.

Література

1. J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.
2. V. Slyusar, M. Protsenko, A. Chernukha, V. Melkin, O. Petrova, M. Kravtsov, S. Velma, N. Kosenko, O. Sydorenko and M. Sobol, “Improving a neural network model for semantic segmentation of images of monitored objects in aerial photographs,”

EasternEuropean Journal of Enterprise Technologies, 2021, vol. 2, no. 6 (114), pp. 86-95. DOI: 10.15587/1729-4061.2021.248390

3. Vadym Slyusar, Ihor Sliusar, Nataliia Bihun, and Volodymyr Piliuhin. Segmentation of analogue meter readings using neural networks. // 44th International Workshop on Modern Machine Learning Technologies and Data Science MOMLET&DS2022, Leiden - Lviv, The Netherlands - Ukraine, 23 -25 Nov. 2022. – Pp. 165 - 175.

4. O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597, 2015. DOI: 10.48550/arXiv.1505.04597.

5. W. Jwaid, Z. Al-Husseini and A. Sabry, Development of brain tumor segmentation of magnetic resonance imaging (MRI) using U-Net deep learning, Eastern-European Journal of Enterprise Technologies, vol. 4, no. 9 (112), 2021, pp. 23-31. doi:10.15587/1729-4061.2021.238957.

6. N. Singh and K. Nongmeikapam, Semantic segmentation of satellite images using deep-UNet, Arabian Journal for Science and Engineering, 2022, pp. 1-13.

7. A. Soni, R. Koner, and V. Villuri, M-Unet: Modified U-Net segmentation framework with satellite imagery, in: Proceedings of the Global AI Congress 2019, Springer, 2020, pp. 47-59.

8. E. Irwansyah, Y. Heryadi, and A. Gunawan, Semantic image segmentation for building detection in urban area with aerial photograph image using U-Net models, in: Proceedings of the 2020 IEEE Asia-Pacific Conf. on Geoscience, Electronics and Remote Sensing Technology (AGERS), 2020, pp. 48-51.

9. Vadym Slyusar, Ihor Sliusar, Pavlenko Anatolii. Improved PSP and U-Net architectures for forest segmentation in remote sensing pictures.// IEEE 2nd Ukrainian Microwave Week (Virtual Event) IEEE UkrMW-2022, 14 – 18 November, 2022. - V. N. Karazin Kharkiv National University, Kharkiv, Ukraine. – 4 p. DOI: 10.1109/UkrMW58013.2022.10037105.

10. N. Singh and K. Nongmeikapam, "Semantic segmentation of satellite images using deep-UNet," Arabian Journal for Science and Engineering, 2022, pp. 1-13.

11. A. Soni, R. Koner, and V. G. K. Villuri, "M-Unet: Modified U-Net segmentation framework with satellite imagery," in Proceedings of the Global AI Congress 2019, Springer, 2020, pp. 47-59.

12. E. Irwansyah, Y. Heryadi, and A. Gunawan, "Semantic image segmentation for building detection in urban area with aerial photograph image using U-Net models," in 2020 IEEE Asia-Pacific Conf. on Geoscience, Electronics and Remote Sensing Technology (AGERS), 2020, pp. 48-51.

13. V. Slyusar. The text segmentation by neural networks of image segmentation.// Artificial Intelligence, 2024, Vol. №98, No 1.- Pp. 46 - 55. DOI: 10.15407/jai2024.01.046.

14. A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998–6008. DOI: 10.48550/arXiv.1706.03762.

15. Слюсар, В.І. Застосування нейромережевих технологій для виявлення підводних боєприпасів.// Вісті вищих учбових закладів. Радіоелектроніка. - Т. 65, № 12, 2022. - 12 с. DOI: 10.20535/S0021347023030020.

16. C. C. Aggarwal, Neural Networks and Deep Learning, 2nd ed. Cham, Switzerland: Springer, 2023. DOI: 10.1007/978-3-031-29642-0.

17. Слюсар В.И., Слюсарь И.И. Анализ эффективности механизмов внимания в задаче сегментации изображений циферблатов аналоговых счетчиков. // Нейромережні технології та їх застосування НМТіЗ-2022: збірник наукових праць XXI-ї Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2022», 7-8 грудня 2022 року, Краматорськ-Тернопіль, Україна. - 7 с.

18. Слюсар В.И. Семейство торцевых произведений матриц и его свойства// Кибернетика и системный анализ. – 1999.- Том 35; № 3.- С. 379-384.- DOI: 10.1007/BF02733426.

19. Слюсар В.І. Архітектурно-математичні основи удосконалення нейронних мереж з класифікації зображень. // Штучний інтелект, 2022, №1.- С. 245 - 258. DOI: 10.15407/jai2022.01.245.

References

1. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

2. V. Slyusar, M. Protsenko, A. Chernukha, V. Melkin, O. Petrova, M. Kravtsov, S. Velma, N. Kosenko, O. Sydorenko and M. Sobol, "Improving a neural network model for semantic segmentation of images of monitored objects in aerial photographs," EasternEuropean Journal of Enterprise Technologies, 2021, vol. 2, no. 6 (114), pp. 86-95. DOI: 10.15587/1729-4061.2021.248390

3. Vadym Slyusar, Ihor Sliusar, Nataliia Bihun, and Volodymyr Piliuhin. Segmentation of analogue meter readings using neural networks. // 44th International Workshop on Modern Machine Learning Technologies and Data Science MOMLET&DS2022, Leiden - Lviv, The Netherlands - Ukraine, 23 -25 Nov. 2022, Pp. 165 - 175.

4. O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597, 2015. DOI: 10.48550/arXiv.1505.04597.

5. W. Jwaid, Z. Al-Husseini and A. Sabry, Development of brain tumor segmentation of magnetic resonance imaging (MRI) using U-Net deep learning, Eastern-European Journal of Enterprise Technologies, vol. 4, no. 9 (112), 2021, pp. 23-31. doi:10.15587/1729-4061.2021.238957.

6. N. Singh and K. Nongmeikapam, Semantic segmentation of satellite images using deep-UNet, Arabian Journal for Science and Engineering, 2022, pp. 1-13.

7. A. Soni, R. Koner, and V. Villuri, M-Unet: Modified U-Net segmentation framework with satellite imagery, in: Proceedings of the Global AI Congress 2019, Springer, 2020, pp. 47-59.

8. E. Irwansyah, Y. Heryadi, and A. Gunawan, Semantic image segmentation for building detection in urban area with aerial photograph image using U-Net models, in: Proceedings of the 2020 IEEE Asia-Pacific Conf. on Geoscience, Electronics and Remote Sensing Technology (AGERS), 2020, pp. 48-51.

9. Vadym Slyusar, Ihor Sliusar, Pavlenko Anatolii. Improved PSP and U-Net architectures for forest segmentation in remote sensing pictures.// IEEE 2nd Ukrainian Microwave Week (Virtual Event) IEEE UkrMW-2022, 14 – 18 November, 2022. - V. N. Karazin Kharkiv National University, Kharkiv, Ukraine. – 4 p. DOI: 10.1109/UkrMW58013.2022.10037105.

10. N. Singh and K. Nongmeikapam, “Semantic segmentation of satellite images using deep-UNet,” Arabian Journal for Science and Engineering, 2022, pp. 1-13.

11. A. Soni, R. Koner, and V. G. K. Villuri, “M-Unet: Modified U-Net segmentation framework with satellite imagery,” in Proceedings of the Global AI Congress 2019, Springer, 2020, pp. 47-59.

12. E. Irwansyah, Y. Heryadi, and A. Gunawan, “Semantic image segmentation for building detection in urban area with aerial photograph image using U-Net models,” in 2020 IEEE Asia-Pacific Conf. on Geoscience, Electronics and Remote Sensing Technology (AGERS), 2020, pp. 48-51.

13. V. Slyusar. The text segmentation by neural networks of image segmentation.//Artificial Intelligence, 2024, Vol. №98, No 1.- Pp. 46 - 55. DOI: 10.15407/jai2024.01.046.

14. A. Vaswani et al., “Attention Is All You Need,” in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998–6008. DOI: 10.48550/arXiv.1706.03762.

15. Vadym Slyusar. Application of neural network technologies for underwater munitions detection. // Radioelectronics and Communications Systems, Vol. 65, No. 12 (2022). - Pp. 654–664. DOI: 10.3103/S0735272723030020..

16. C. C. Aggarwal, Neural Networks and Deep Learning, 2nd ed. Cham, Switzerland: Springer, 2023. DOI: 10.1007/978-3-031-29642-0.

17. Slyusar V.I., Sliusar I.I. Analysis of the efficiency of attention mechanisms in the task of segmenting images of analog meter dials. // 21st International Scientific Conference "Neural Network Technologies and Applications NNTA-2022", December 7-8, 2022. Kramatorsk-Ternopil, Ukraine. - 7 p.

18. Slyusar, V. I. A Family of Face Products of Matrices and its Properties. // Cybernetics and Systems Analysis c/c of Kibernetika i Sistemnyi Analiz. – 1999, Vol 35; Part 3, pages 379-384. – Consultants Bureau (USA). - DOI: 10.1007/BF02733426.

19. Slyusar, V. I. Architectural and mathematical fundamentals of improvement neural networks for classification of images. // Artificial Intelligence, 2022, №1.- Pp. 245 - 258. DOI: 10.15407/jai2022.01.245.

The article has been sent to the editors 17.11.24.
After processing 23.11.24.
Submitted for printing 30.12.24.